

Identification of protein pockets and cavities by Euclidean Distance Transform

Sebastian Daberdaku

Department of Information Engineering, University of Padova, via Gradenigo 6/A, 35131 Padova (PD), Italy

Email address: sebastian.daberdaku@unipd.it

Abstract

Protein pockets and cavities usually coincide with the active sites of biological processes, and their identification is significant since it constitutes an important step for structure-based drug design (Y. Lee et al. 2018) and protein-ligand docking applications (Krivák et al. 2015). This paper presents a novel purely geometric algorithm for the detection of ligand binding protein pockets and cavities based on the Euclidean Distance Transform (EDT). The EDT can be used to compute the Solvent-Excluded surface for any given probe sphere radius value at high resolutions and in a timely manner (Daberdaku et al. 2016, 2018). The algorithm is adaptive to the specific candidate ligand: it computes two voxelised protein surfaces using two different probe sphere radii depending on the shape of the candidate ligand. The pocket regions consist of the voxels located between the two surfaces, which exhibit a certain minimum depth value from the outer surface. The distance map values computed by the EDT algorithm during the second surface computation can be used to elegantly determine the depth of each candidate pocket and to rank them accordingly. Cavities on the other hand, are identified by scanning the inside of the protein for voids. The algorithm determines and outputs the best k candidate pockets and cavities, i.e. the ones that are more likely to bind to the given ligand. The method was applied to a representative set of protein-ligand complexes and their corresponding unbound protein structures to evaluate its ligand binding site prediction capabilities, and was shown to outperform most of the previously developed purely geometric pocket and cavity search methods.

References

- Daberdaku S, Ferrari C. (2016). Computing Discrete Fine-Grained Representations of Protein Surfaces. *LNBI Computational Intelligence Methods for Bioinformatics and Biostatistics: 12th International Meeting, Revised Selected Papers, Lecture Notes in Bioinformatics* 9874: 180–195.
- Daberdaku S, Ferrari C. (2018). Computing Voxelised Representations of Macromolecular Surfaces: A Parallel Approach. *The International Journal of High Performance Computing Applications* 32 (3): 407–32.
- Krivák R, Hoksza D. (2015). Improving Protein-Ligand Binding Site Prediction Accuracy by Classification of Inner Pocket Points Using Local Features. *Journal of Cheminformatics* 7 (1): 12.
- Lee Y, Basith S, Choi S. (2018). Recent Advances in Structure-Based Drug Design Targeting Class A G Protein-Coupled Receptors Utilizing Crystal Structures and Computational Simulations. *Journal of Medicinal Chemistry* 61 (1): 1–46.